In [5]:
```
## import all the required packages.
import os
import numpy as np
from os import listdir
from imageio import imread
from keras.utils import to_categorical
from sklearn.model_selection import train_test_split
from keras.utils.image_utils import img_to_array

import PIL
import matplotlib.pyplot as plt
```

In [6]:
```
# Settings
num_classes = 10
test_size = 0.2
```

This function is used to read the picture from the `data_path` and convert the picture to black and white

In [7]:
```
def get_img(data_path):
  ## Getting image array from path:
  img = PIL.Image.open(data_path)
  img = img.convert("L")
  img = img_to_array(img)
  img = np.resize(img, (100, 100, 1))
  return img
```

Get dataset from picture and then split to train and test set

In [8]:
```
dataset_path = "/content/drive/MyDrive/HUDK_4050_Final/Dataset"

## Getting all data from data path
labels = sorted(listdir(dataset_path)) ## in order to read the files from the sorted li
X = []
Y = []
for i, label in enumerate(labels):
  data_path = dataset_path + "/" + label

  for data in listdir(data_path):
    ## create dataset
    img = get_img(data_path + "/" + data)
    X.append(img) ## X is the source file for all pictures.
    Y.append(i)   ## Y is the number represented for all the picutres.
## transfer X, and Y
X = 1 - np.array(X).astype("float32") /255
Y = np.array(Y).astype("float32")
Y = to_categorical(Y, num_classes)
## split out the dataset
X, X_test, Y, Y_test = train_test_split(X, Y, test_size=test_size, random_state = 42)
print(X.shape)
print(X_test.shape)
print(Y.shape)
print(Y_test.shape)
```

```
(1649, 100, 100, 1)
(413, 100, 100, 1)
(1649, 10)
(413, 10)
```